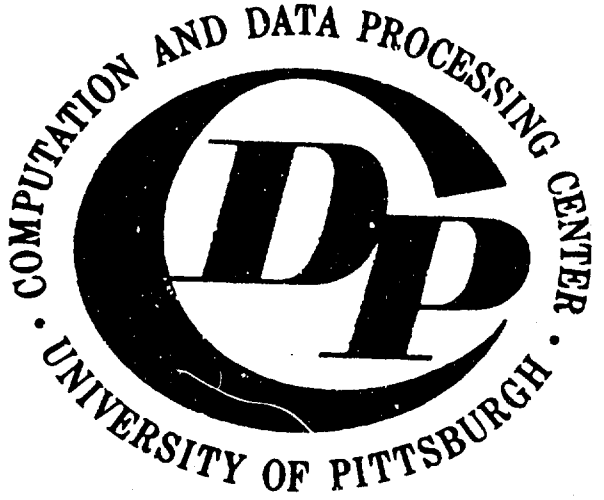

800 CATHEDRAL OF LEARNING · PITTSBURGH, PENNSYLVANIA 15213

UNIVERSITY OF
PITTSBURGH
GENERALIZED
RECORDING
AND
DISSEMINATION
EXPERIMENT

A.R.P.A.
Advanced Research Projects Agency

Conducted by

UNIVERSITY OF PITTSBURGH
COMPUTATION AND DATA PROCESSING CENTER
PITTSBURGH PENNSYLVANIA 15213

Progress in Computerized Typesetting

Lee Ohringer

A reprint of a paper presented before
The 17th Annual Meeting of the
Technical Association of the Graphic Arts
Park Plaza, Toronto, Ontario, Canada
June 1, 1965

Progress in Computerized Typesetting

Lee Ohringer
Project UPGRADE Coordinator

University of Pittsburgh
Computation and Data Processing Center
800 Cathedral of Learning
Pittsburgh, Pennsylvania 15213

Abstract

Recent years have seen the introduction of digital computers into the printing industry. Thus far computers have been used primarily in the accounting departments and for limited typesetting--for example, in newspaper work.

At the University of Pittsburgh we are studying new ways in which the computer can be used to aid the printer. We are experimenting with advanced concepts such as computerized editing routines and typesetting of complex material. The programs which we have written for our computer enable the editor to see the changes he wishes effected immediately on the text using a display screen which is electronically controlled by our computer.

We have also developed computerized indexing methods and have used our computer to generate a dictionary of current scientific terms from the text which we have collected.

Presently two IBM 1401's, an IBM 7070 and 7090, and a PDP-4 computer are available for our research with our Photon 560.

This project receives partial support from the Department of Defense Advanced Research Projects Agency under contract SD-186 and National Science Foundation Grants GP 2310 and G 11309.

INTRODUCTION

This paper covers four areas of research currently being investigated at the University of Pittsburgh Computation and Data Processing Center. The first of these is a project to collect large amounts of text in computer compatible form. The second is a user-oriented computer language which we designed specifically to simplify research on this and other text. Computerized typesetting comprises my third subject, and editing, formating and incorporating author's alteration using computers is the final topic which I will discuss in this paper.

UNIVERSITY OF PITTSBURGH GENERALIZED RECORDING AND DISSEMINATION EXPERIMENT

A.R.P.A.
Advanced Research Projects Agency

Conducted by

UNIVERSITY OF PITTSBURGH
COMPUTATION AND DATA PROCESSING CENTER
PITTSBURGH, PENNSYLVANIA 15213

Our entire efforts in this field fall into the area covered by Project UPGRADE which stands for the University of Pittsburgh Generalized Recording and Dissemination Experiment.

TEXT COLLECTION

It was our text collection project which gave us our initial contact with the printing industry. Since our Computing Center has been primarily devoted to developing methods of text handling and information processing, as contrasted with mathematical methods research done at most computing centers, it was natural that we were the ones requested by the Department of Defense to develop a means to obtain large amounts of text in computer readable form.

Toward this goal, we examined methods used by past projects such as punching the text onto tab cards or paper tape. We also considered optical character readers which were then being proposed. None of these methods appeared to be capable of meeting our needs.

It was then that we turned to the printing industry in search of an answer. We found that many printers were sincerely interested in what we were doing and quite willing to help in any way they could. A pilot study was set up whereby we receive the typesetting tapes from Lancaster Press and from a job that Kingsport Press was doing for McGraw-Hill. Since that time the list of printers, publishers, and research centers from whom we have received advice and co-operation has grown so that today over fifty have contributed significantly to our efforts. The following slide shows many of those to whom we owe credit for much of our success.

```
AD PRESS, LTD.
AMERICAN CHEMICAL SOCIETY
AMERICAN MATHEMATICAL SOCIETY
AMERICAN METEOROLOGICAL SOCIETY
AMERICAN NEWSPAPER PUBLISHERS ASSOC
AMERICAN PRINTING HOUSE FOR THE BLIND
ASSOCIATION OF AMERICAN UNIVERSITY PRESSES
CENTER FOR APPLIED LINGUISTICS
COLONIAL PRESS INC
COMPOSITION INFORMATION SERVICES
COMPUTER TYPESETTING INC
CURTIS PUBLISHING CO
DUQUESNE UNIVERSITY PRESS
FEDERATION OF AMERICAN SOCIETIES FOR EXPERIMENTAL BIOLOGY
GRAPHIC ARTS TECHNICAL FOUNDATION
HARRIS-INTERTYPE CORPORATION
HARVARD UNIVERSITY
HERBICK AND HELD
INDIANA STATE COLLEGE
INFORONICS
INTERNATIONAL BUSINESS MACHINE COPORATION
INTERNATIONAL TYPOGRAPHIC COMPOSITION ASSOC INC
ITEK CORPORATION
JOHN WILEY AND SONS INC
JOHNS HOPKINS UNIVERSITY
KINGSPORT PRESS INC
LANCASTER PRESS
LANSTON MONOTYPE CO
LOS ANGELES TIMES
LOUISIANA STATE UNIVERSITY PRESS
MACK PRINTING CO
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MC GRAW-HILL BOOK COMPANY INC
MERGENTHALER LINOTYPE CO
NATIONAL INSTITUTES OF HEALTH
PHOTON INC
PRINTING PRODUCTION MAGAZINE
RAND CORPORATION
RANDOM HOUSE INCORPORATED
RESEARCH AND ENGINEERING COUNCIL OF THE GRAPHIC ARTS
ROCAPPI
SUMMER INSTITUTE OF LINGUISTICS
SYSTEM DEVELOPMENT CORPORATION
THOMAS NELSON AND SONS
U S GOVERNMENT PRINTING OFFICE
UNIVERSITY OF ALABAMA PRESS
UNIVERSITY OF NORTH CAROLINA
UNIVERSITY OF WISCONSIN PRESS
VAIL-BALLOU PRESS INC
W A BENJAMIN, INC
W B SAUNDERS CO
WAVERLY PRESS
WILLIAMS AND WILKINS PUBLISHERS
```

We are, of course, still interested in making further arrangements to receive the typesetting tapes from other printers. Of particular interest and use to us are the tapes from books by renowned authors, poetry collections, biographies, and versions of the various Bibles. We use the text from these tapes solely for research in computerized text processing such as automatic indexing, abstracting, and classification--never in a way not approved by the printer who supplied it.

One obvious benefit to the printer from our text collection will be the printing needs created for publication of the research done by us and others on the text which we accumulate.

JET	423	GENETIC	423	DEFINITION	423		
SHEAR	422	RAY	422	ANTENNA	422		
SPECIFIED	421	SELECTED	421	METABOLIC	421		
DEVELOPS	420	WHO	419	STRAIN	419		
REAL	419	MOTIONS	419	HORMONES	419		
DEPOSITED	419	SEX	418	RELATIONS	418		
INTEGRAL	418	FURNACE	418	VISCOSITY	417		
PIPE	417	CONTINUOUSLY	417	BOND	417		
DRAWN	416	OCCURRENCE	415	INFRARED	415		
REACT	414	MOSTLY	414	SHIPS	413		
HOLES	413	CHANGING	413	BANDS	413		
ISOTOPES	412	FLOOR	412	FAIRLY	411		
CONTINENTAL	410	WHAT	409	SIX	409		
SUBSEQUENT	408	SEEMS	408	NICKEL	408		
DIAGRAM	408	BONDS	408	SIMPLY	407		
ORDERS	407	GENERATION	406	FERMENTATION	406		
RELATIONSHIPS	405	HIGHEST	405	VISUAL	404		
CONE	403	ACTS	403	TRANSPORT	402		
OBSERVATION	402	MICROORGANISMS	402	FAILURE	402		
CROP	402	CELLULOSE	402	ALLOY	402		
LIMESTONE	401	COLORS	401	BRANCH	401		
SCATTERING	400	LONGITUDINAL	399	LEGS	399		
ENVIRONMENTAL	399	DORSAL	399	ATTACK	399		
DEGREES	398	CONFIGURATION	398	OPERATED	397		
APPARATUS	397	SEEDS	396	ORBIT	396		
EVAPORATION	396	COMBINATIONS	396	TESTING	395		
PACIFIC	395	ALONE	395	PROCEDURES	394		
OSCILLATOR	394	SULFATE	393	PHYSIOLOGICAL	393		
CAUSING	393	BRIDGE	393	VARIABLES	392		
PLACES	392	PARTLY	392	INSTANCE	392		
ATLANTIC	392	TREATED	391	RADIAL	391		
TAIL	390	PROPAGATION	390	TENDENCY	389		
SATISFACTORY	389	STEMS	388	FRUITS	388		
STATIONS	387	SITE	387	GIVING	387		
GRADIENT	386	TRANSITION	385	SENSORY	385		
IDENTICAL	385	DRIVE	385	DISCOVERED	385		
PRINTING	384	OXIDES	384	DEPTHS	384		
SPECTRA	383	PARTIALLY	383	MATHEMATICAL	383		
ADAPTED	383	REPTILES	382	CALIFORNIA	382		

This slide shows such an example from a page of the descending frequency list from the McGraw-Hill Encyclopedia of Science and Technology. This listing was created completely automatically by our computer from the text which we collected from the Teletypesetter tapes used to do the printing. Also, we can make statistics available to the printer from the tapes he supplies. Such statistics could be useful in helping him design a new matrix arrangement, for example.

PENELOPE (Pitt Natural Language Processor)

PENELOPE, the Pitt Natural Language Processor, was designed to satisfy the need for a computer language capable of processing text efficiently and easily. PENELOPE was designed specifically to allow the

programmer to write his program in a way which would be natural to him. PENELOPE then translates his statements into code which can be understood and executed by a computer. Examples of PENELOPE's capabilities are shown and explained in a paper which I presented at last year's TAGA meeting in Pittsburgh. This paper appeared in its complete form in the 1964 TAGA proceeding, therefore I will not go into detail here.

The translator for PENELOPE has been completed and is in use on the IBM 7070 at the University of Pittsburgh. Copies of this program are available, free, upon request, as are most of the routines developed by our Center. A technical write-up is also available upon request.

COMPUTERIZED TYPESETTING

Our progress in computerized typesetting, since my talk at last year's meeting, involves our advancing from a theoretical approach to actual production. Last year I spoke of what could be done if we had a piece of photo-typesetting equipment. This year I will tell you what we have done with the Photon-560 which we have since acquired and what we are planning to do.

For the justification part of our system we are using a modified version of the PC6 system which was originally conceived by Dr. Michael P. Barnett. One feature of the original system which we hoped to improve was to reduce the great number of keystrokes required to insert the printing control information such as type size and type font.

We feel we have accomplished a means of doing this as we demonstrated when we prepared the control tapes for a bibliography for learning research as shown on my next slide.

1 108. Traxler, Arthur E.
2 "One Reading Test Serves the Purpose,"
3 [The Clearing House,] XIV (March, 1940), 419-21.
4 Presents correlations (a) between scores on different
forms of four reading tests administered a
year apart and (b) between different reading tests
administere: at intervals of one year.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

[AL,ML,"1][CP,DL5]Traxler, Arthur E.
[DL2]""One Reading Test Serves the Purpose,''
[DL4][DL16]The Clearing House,[DL4] XIV (March, 1940), 419-21.
[DL2]Presents correlations (a) between scores on
different forms of four reading tests administered
a year apart and (b) between different reading
tests administered at intervals of one year.

In punching these tapes the only signals to the computer which the keyboarder inserted were the code numbers 1, 2, 3, 4 in the left hand margin and brackets around any text which was to be in italics. With a simple pre-processing computer program we then expanded these into the appropriate codes, thereby eliminating many key-strokes. This slide shows one of the entries from this bibliography. The top of the slide shows how it appeared as originally keyboarded and below is shown how it looked after the control codes were automatically inserted.

Another feature which I indicated that we were going to add to our computer-typesetting system was the hyphen-

ation capability. Currently, we have a member of our staff working on such a routine and hope to have it finished by the end of the summer. However, even after our hyphenation routine is completed our computer will try to justify each line by word spacing and letter spacing, as we have been doing, in order to save computer time. In an effort to maintain graphic arts quality we have set upper and lower limits on such spacing.

EDITING, FORMATING AND AUTHOR'S ALTERATIONS

It is in the area of man-machine editing that I feel we have made our most significant progress. We have written and are currently using a general purpose text editing/formating routine. This program is written for a small scale computer (the PDP-4) which is connected to our 7090 on an interrupt basis. Text can be accepted either from cards, magnetic tape, or the various kinds of paper tape. The text is then displayed on a cathode ray tube screen, and the operator is able to make the changes he desires by use of a light pen and a typewriter keyboard.

The operator can use the light pen to indicate which of several editing functions he wishes to perform. He does this simply by pointing his light pen at the desired function which appears at the bottom of the screen. A picture of the screen containing these codes is shown on my next slide.

+The MOV function:

The words poorly placed are rearranged for better style.+

```
RMT WMT DMT WTM RWD SBC
TYP TYH CLR DEL MOV SPC   IN OUT DMP BIG
#HS SLO FWD REV RUN HLT MAN   CO   LO ORD
```

Currently the editing program has the ability to

(1) RMT (Read Magnetic Tape) - Read input or corrections from magnetic tape.

(2) WMT (Write Magnetic Tape) - Copy the text which is currently on the screen onto magnetic tape. (Does not alter what is on screen.)

(3) DMT (Dump Magnetic Tape) - Write the text which is on the screen onto magnetic tape and clear the screen.

(4) WTM (Write Tape Mark) - End of current job.

(5) RWD (Rewind Magnetic Tapes) - Go to the beginning of the magnetic tapes.

(6) SBC (Switch B and C) - Interchange the input and output tapes to allow the user to read back what he has just written.

(7) TYP (TYPe) - This will produce on the typewriter a hard copy of the contents of the screen.

(8) TYH (TYpe Halt) - This command will stop the typing.

(9) CLR (CLeaR) - Erase the text from the screen.

(10) DEL (DELete) - Erase a specified part of the text.

(11) MOV (MOVe) - Move a specified part of the text to another specific point.

(12) SPG (Special Pattern Generator) - This control allows the user to change the character set being used.

(13) IN (IN) - Read paper tape, display text on screen.

(14) OUT (OUT) - Punch paper tape containing text from the screen but leave text on screen.

(15) DMP (DuMP) - Punch paper tape containing the text from the screen and clear the screen.

(16) BIG (BIG) - Punch paper tape so that the holes form the shapes of the letters on the screen.

(17) RUN (RUN) - This light button will cause the text to move up the screen with the first line disappearing off the top and additional text appearing along the bottom.

(18) FAS (FASt) - This will cause the text to move faster (See RUN).

(19) SLO (SLOw) - This will cause the text to move slower.

(20) FWD (ForWarD) - This will cause the text to move up the screen and is used to cancel the affect of

the REV command.

(21) REV (REVerse) - This will cause the text to backup with the top lines reappearing and the bottom lines disappearing.

(22) HLT (HaLT) - This will stop the text from moving.

(23) MAN (MANual) - This command will move the text one line at a time in same way as RUN.

As I have indicated some commands such as MOV work only with a specified portion of the text. The last three light button allow pointers to be placed in the text to specify what is to be moved.

(24) LD (Left Delimeter) - will allow placement of the left pointer.

(25) RD (Right Delimeter) - will allow placement of the right pointer, and

(26) CD (Cursor Defined) - will allow placement of an additional pointer to indicate to where the text is to be moved.

Presently all of these commands are built in only through programming and are not part of the hardware. This allows us a great amount of flexibility in making modifications and additions. For example, one addition which is currently being considered in the COPY command which will allow the operator to duplicate some portion of the text on the screen. Another alteration which we are considering is to divide the screen in half, by programming of course, in order to be able to accept and output text from two independent sources. Then the main text could

be read into the top half of the screen and insertions could be read into the bottom. The operator could combine them as he wishes.

As a testimonial to the usability of these routines, several of the secretaries on our staff, with absolutely no computer training have used this routine in typing papers in order to allow for ease of "author alterations." In fact, the preliminary drafts of the paper I have just presented were prepared using this system.

Bibliography

Bacon, Charles R. T., Text Editor 2, Technical Report, University of Pittsburgh Computation and Data Processing Center, Pittsburgh, Pennsylvania April 8, 1965

Barnett, M. P., and D. A. Luce, The TYPRINT System Cooperative Computing Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, Reprinted May 3, 1965 by the University of Pittsburgh Computation and Data Processing Center.

Isner, Dale W., PENELOPE, The Pitt Natural Language Processor, University of Pittsburgh Computation and Data Processing Center, Pittsburgh, Pennsylvania, April, 1965.

Ohringer, Lee, "Computer Input from Printing Control Tapes," TAGA Proceedings 1964, p. 304-316, Rochester, New York.